MULTIPLE APPROACHES TO ROBUST SPEECH RECOGNITION

*Richard M. Stern, Fu-Hua Liu, Yoshiaki Ohshima, Thomas M. Sullivan, Alejandro Acero**

Department of Electrical and Computer Engineering
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

This paper compares several different approaches to robust speech recognition. We review CMU's ongoing research in the use of acoustical pre-processing to achieve robust speech recognition, and we present the results of the first evaluation of pre-processing in the context of the DARPA standard ATIS domain for spoken language systems. We also describe and compare the effectiveness of three complementary methods of signal processing for robust speech recognition: acoustical pre-processing, microphone array processing, and the use of physiologically-motivated models of peripheral signal processing. Recognition error rates are presented using these three approaches in isolation and in combination with each other for the speaker-independent continuous alphanumeric census speech recognition task.

1. INTRODUCTION

The need for speech recognition systems and spoken language systems to be robust with respect to their acoustical environment has become more widely appreciated in recent years (*e.g.* [1]).

Results of several studies have demonstrated that even automatic speech recognition systems that are designed to be speaker independent can perform very poorly when they are tested using a different type of microphone or acoustical environment from the one with which they were trained (*e.g.* [2, 3]), even in a relatively quiet office environment. Applications such as speech recognition over telephones, in automobiles, on a factory floor, or outdoors demand an even greater degree of environmental robustness.

The CMU speech group is committed to the development of speech recognition systems that are robust with respect to environmental variation, just as it has been an early proponent of speaker-independent recognition. While most of our work presented to date has described new *acoustical pre-processing* algorithms (*e.g.* [2, 4, 5], we have always regarded pre-processing as one of several approaches that must be developed in concert to achieve robust recognition.

The purpose of this paper is twofold. First, we describe our results for the DARPA benchmark evaluation for robust speech recognition for the ATIS task, discussing the effectiveness of our methods of acoustical pre-preprocessing in the context of this task. Second, we describe and compare the effectiveness of three complementary methods of signal processing for robust speech recognition: acoustical pre-processing, microphone array processing, and the use of physiologically-motivated models of peripheral signal processing.

2. ACOUSTICAL PRE-PROCESSING

We have found that two major factors degrading the performance of speech recognition systems using desktop microphones in normal office environments are additive noise and unknown linear filtering. We showed in [2] that simultaneous *joint* compensation for the effects of additive noise and linear filtering is needed to achieve maximal robustness with respect to acoustical differences between the training and testing environments of a speech recognition system. We described in [2] two algorithms that can perform such joint compensation, based on additive corrections to the cepstral coefficients of the speech waveform.

The first compensation algorithm, *SNR-Dependent Cepstral Normalization* (SDCN), applies an additive correction in the cepstral domain that depends exclusively on the instantaneous SNR of the signal. This correction vector equals the average difference in cepstra between simultaneous "stereo" recordings of speech samples from both the training and testing environments at each SNR of speech in the testing environment. At high SNRs, this correction vector primarily compensates for differences in spectral tilt between the training and testing environments (in a manner similar to the blind deconvolution procedure first proposed by Stockham *et al.* [6]), while at low SNRs the vector provides a form of noise subtraction (in a manner similar to the spectral subtraction algorithm first proposed by Boll [7]). The SDCN algorithm is simple and effective, but it requires environment-specific training.

The second compensation algorithm, *Codeword-Dependent Cepstral Normalization* (CDCN), uses EM techniques to compute ML estimates of the parameters characterizing the contributions of additive noise and linear filtering that when applied in inverse fashion to the

*Present address: Telefónica Investigación y Desarrollo, Emilio Vargas 6, Madrid 28043, Spain

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1992		**00-00-1992 to 00-00-1992**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multiple Approaches to Robust Speech Recognition	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Carnegie Mellon University,School of Computer Science,Pittsburgh,PA,15213**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**6**	

cepstra of an incoming utterance produce an ensemble of cepstral coefficients that best match (in the ML sense) the cepstral coefficients of the incoming speech in the testing environment to the locations of VQ codewords in the training environment. The CDCN algorithm has the advantage that it does not require *a priori* knowledge of the testing environment (in the form of stereo training data in the training and testing environments), but it is much more computationally demanding than the SDCN algorithm. Compared to the SDCN algorithm, the CDCN algorithm uses a greater amount of structural knowledge about the nature of the degradations to the speech signal in order to achieve good recognition accuracy. The SDCN algorithm, on the other hand, derives its compensation vectors entirely from empirical observations of differences between data obtained from the training and testing environments.

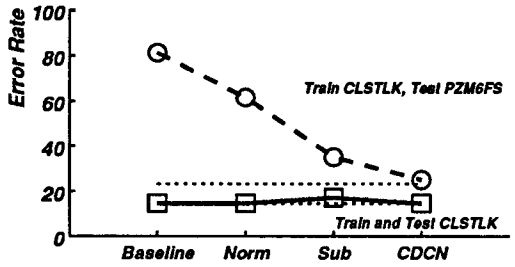

Figure 1: Comparison of error rates obtained on the census task with no processing, spectral subtraction, spectral normalization, and the CDCN algorithm. SPHINX was trained on the CLSTLK microphone and tested using either the CLSTLK microphone (solid curve) or the PZM6FS microphone (broken curve).

Figure 1 compares the error rate obtained when the SPHINX system is trained using the DARPA standard HMD-414 closetalking microphone (CLSTLK), and tested using either the CLSTLK microphone or the omnidirectional desktop Crown PZM-6FS microphone (PZM6FS). The census database was used, which contains simultaneous recoredings of speech from the CLSTLK and PZM6FS microphones in the context of a speaker-independent continuous-speech alphanumeric task with perplexity 65 [2]. These results demonstrate the value of the joint compensation provided by the CDCN algorithm in contrast to the independent compensation using either spectral subtraction or spectral normalization. The horizontal dotted lines indicate the recognition accuracy obtained when the system is tested on the microphone with which it was trained, with no processing. The intersection of the upper curve with the upper horizontal line indicates that with CDCN compensation, SPHINX can recognize speech using the PZM6FS microphone just as well when trained on the CLSTLK microphone as when trained using the PZM6FS.

More recently we have been attempting to develop new algorithms which combine the computational simplicity of SDCN with the environmental independence of CDCN. One such algorithm, *Blind SNR-Dependent Cepstral Normalization* (BSDCN) avoids the need for environment-specific training by establishing a correspondence between

ALGORITHM	ENVIRN. SPEC?	COMPLEXITY	ERR RATE
NONE	NO	NONE	68.6%
SDCN	YES	MINIMAL	27.6%
CDCN	NO	GREATER	24.3%
BSDCN	NO	MINIMAL	30.0%

Table 1: Comparison of recognition accuracy of SPHINX with no processing and the CDCN, SDCN, and BSDCN algorithms. The system was trained using the CLSTLK microphone and tested using the PZM6FS microphone. Training and testing on the CLSTLK produces a recognition accuracy of 86.9%, while training and testing on the PZM6FS produces 76.2%

SNRs in the training and testing environments by use of traditional nonlinear warping techniques [8] on histograms of SNRs from each of the two environments [5]. Table 1 compares the environmental specificity, computational complexity, and recognition accuracy of these algorithms when evaluated on the alphanumeric database described in [2]. Recognition accuracy is somewhat different from the figures reported in Fig. 1 because the version of SPHINX used to produce these data was different. All of these algorithms are similar in function to other currently-popular compensation strategies (*e.g.* [3, 9]).

The DARPA ATIS robust speech evaluation. The original CDCN algorithm described in [2] was used for the February, 1992, ATIS-domain robust-speech evaluation. For this evaluation, the SPHINX-II system was trained using the CLSTLK microphone, and tested using both the CLSTLK microphone and the unidirectional Crown PCC-160 microphone (PCC160). All incoming speech in this evaluation was processed by the CDCN algorithm, regardless of whether the testing environment was actually the CLSTLK or PCC160 microphone, and the CDCN algorithm was not provided with explicit knowledge of the identity of the environment within which it is operating.

As described elsewhere in these Proceedings [10] , the system used for the official robust-speech evaluations was not trained as thoroughly as the baseline system was trained. Specifically, the official evaluations were performed after only a single iteration through training data that was processed with the CDCN algorithm, and without the benefit of general English sentences in the training database.

In Fig. 2 we show the results of an unofficial evaluation of the SPHINX-II system that was performed immediately after the official evaluation was complete. The purpose of this second evaluation was to evaluate the improvement provided by an additional round of training with speech processed by CDCN, in order to be able to directly compare error rates on the ATIS task with CDCN with those produced by a comparably-trained system on the same data, but without CDCN. As Fig. 2 shows, using the

CDCN algorithm causes the error rate to increase from 15.1% to only 20.4% as the testing microphone is changed from the CLSTLK to the PCC160 microphone. In contrast, the error rate increases from 12.2% to 38.8% when one switches from the CLSTLK to the PCC160 microphone without CDCN.

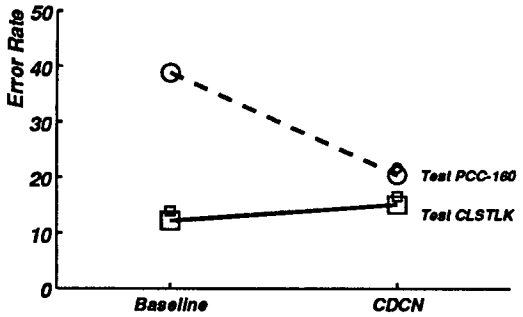

Figure 2: Comparison of error rates obtained on the DARPA ATIS task with no processing, spectral subtraction, spectral normalization, and the CDCN algorithm. SPHINX-II was trained on the CLSTLK microphone in all cases, and tested using either the CLSTLK microphone (solid curve) or the cardiod desktop Crown PCC160 microphone (broken curve).

Only two sites submitted data for the present robust speech evaluation. CMU's percentage degradation in error rate in changing from the CLSTLK to the PCC160 environment, as well as the absolute error rate obtained using the PCC160 microphone, were the better of the results from these two sites.

3. MICROPHONE ARRAYS AND ACOUSTICAL PRE-PROCESSING

Despite the encouraging results that we have achieved using acoustical pre-processing, we believe that further improvements in recognition accuracy can be obtained in difficult environments by combining acoustical pre-processing with other complementary types of signal processing. The use of microphone arrays is motivated by a desire to improve the effective SNR of speech as it is input to the recognition system. For example, the headset-mounted CLSTLK microphone produces a higher SNR than the PZM6FS microphone under normal circumstances because it picks up a relatively small amount of additive noise, and the incoming signal is not degraded by reverberated components of the original speech.

To estimate the potential significance of the reduced SNR provided by the PZM6FS microphone in the office environment, we manually examined all utterances in the test set of the census task that were recognized correctly when training and testing with the CLSTLK microphone but that were recognized incorrectly when training and testing using the PZM6FS. We found that 54.7 percent of these errors were caused by the confusion of silence or noise segments with weak phonetic events, (20 percent of the errors were caused by cross-talk from other noise sources in the room, and the remaining errors could not be attributed to a particular cause.) Microphone arrays can, in principle, produce directionally-sensitive gain patterns that can be adjusted to produce maximal sensitivity in the direction of the speaker and reduced sensitivity in the direction of competing sound sources. To the extent that such processing could improve the effective SNR at the input to a speech recognition system, the error rate would be likely to be substantially decreased, because the number of confusions between weak phonetic events and noise would be sharply reduced.

Several different types of array-processing strategies have been applied to automatic speech recognition. The simplest approach is that of the delay-and-sum beamformer, in which delays are inserted in each channel to compensate for differences in travel time between the desired sound source and the various sensors (*e.g.* [11, 12]). A second option is to use an adaptation algorithm based on minimizing mean square energy such as the Frost or Griffiths-Jim algorithm [13]. These algorithms provide the opportunity to develop nulls in the direction of noise sources as well as more sharply focused beam patterns, but they assume that the desired signal is statistically independent of all sources of degradation. Consequently, these algorithms can provide good improvement in SNR when signal degradations are caused by additive independent noise sources, but these algorithms do not perform well in reverberant environments when the distortion is at least in part a delayed version of the desired speech signal [14, 15]. (This problem can be avoided by only adapting during non-speech segments [16]). A third type of approach to microphone array processing is to use a cross-correlation-based algorithm that isolates inter-sensor differences in arrival time of the signals directly (*e.g.* [17]). These algorithms are appealing because they are based on human binaural hearing, and cross-correlation is an efficient way to identify the direction of a strong signal source. Nevertheless, the nonlinear nature of the cross-correlation operation renders it inappropriate as a means to directly process waveforms. We believe that signal processing techniques based on human binaural perception are worth pursuing, but their effectiveness for automatic speech recognition remains to be conclusively demonstrated.

Pilot evaluation of the Flanagan array. In order to obtain a better understanding of the ability of array processing to provide further improvements in recognition accuracy we conducted a pilot evaluation of the 23-microphone array developed by Flanagan and his colleagues at AT&T Bell Laboratories. The Flanagan array, which is described in detail in [11, 12], is a one-dimensional delay-and-sum beamformer which uses 23 microphones that are unevenly spaced in order to provide a beamwidth that is approximately constant over the range of frequencies of interest. The array uses first-order gradient microphones, which develop a null response in the vertical plane. We wished to compare the recognition accuracy on the census task obtained using the Flanagan array with the accuracy observed using the CLSTLK and PZM6FS

microphones. We were especially interested in determining the extent to which array processing provides an improvement in recognition accuracy that is complementary to the improvement in accuracy provided by acoustical pre-processing algorithms such as the CDCN algorithm.

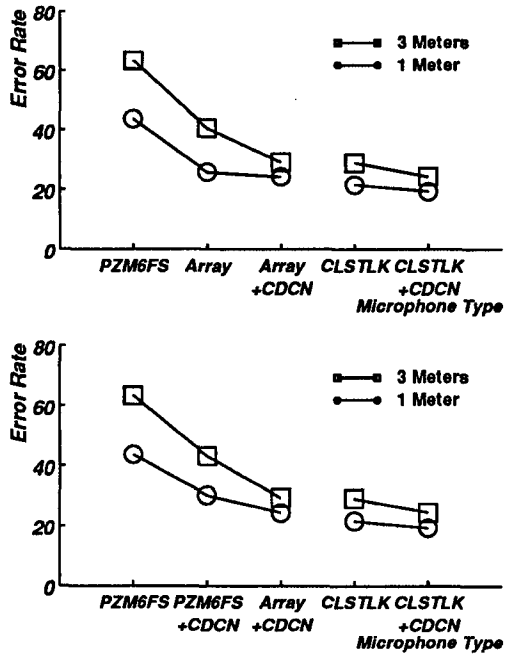

Figure 3: Comparison of recognition accuracy obtained on a portion of the census task using the omnidirectional Crown PZM-6FS, the 23-microphone array developed by Flanagan, and the Senneheiser microphone, each with and without CDCN. Data were obtained from simultaneous recordings using the three microphones at distances of 1 and 3 meters (for the PZM-6FS and the array).

14 utterances from the census database were obtained from each of five male speakers in a sparsely-furnished laboratory at the Rutgers CAIP Center with hard walls and floors. The reverberation time of this room was informally estimated to be between 500 and 750 ms. Simultaneous recordings were made of each utterance using three microphones: the Sennheiser HMD-414 (CLSTLK) microphone, the Crown PZM6FS, and the Flanagan array with input lowpass-filtered at 8 kHz. Recordings were made with the speaker seated at distances of 1, 2, and 3 meters from the PZM6FS and Flanagan array microphones, wearing the CLSTLK microphone in the usual fashion at all times.

Figure 3 summarizes the error rates obtained from these speech samples at two distances, 1 and 3 meters, with and without the CDCN algorithm applied to the output of the microphone array. Error rates using the CLSTLK microphone differed somewhat for the two distances because different speech samples were obtained at each distance and because the sample size is small. The SPHINX system had been previously trained on speech obtained using the CLSTLK microphone. As expected, the worst results were obtained using the PZM6FS microphone, while the lowest error rate was obtained for speech recorded using the CLSTLK. More interestingly, the results in Fig. 3 show that both the Flanagan array and the CDCN algorithm are effective in reducing the error rate, and that in fact the error rate at each distance obtained with the combination of the two is very close to the error rate obtained with the CLSTLK microphone and no acoustical pre-processing. The complementary nature of the improvement of the Flanagan array and the CDCN algorithm is indicated by the fact that adding CDCN to the array improves the error rate (upper panel of Fig. 3), and that converting to the array even when CDCN is already employed also improves performance (lower panel).

4. PHYSIOLOGICALLY-MOTIVATED FRONT ENDS AND ACOUSTICAL PRE-PROCESSING

In recent years there has also been an increased interest in the use of peripheral signal processing schemes that are motivated by human auditory physiology and perception, and a number of such schemes have been proposed (*e.g.* [18, 19, 20, 21]). Recent evaluations indicate that with "clean" speech, such approaches tend to provide recognition accuracy that is comparable to that obtained with conventional LPC-based or DFT-based signal processing schemes, but that these auditory models can provide greater robustness with respect to enviromental changes when the quality of the incoming speech (or the extent to which it resembles speech used in training the system) decreases [22, 23]. Despite the apparent utility of such processing schemes, no one has a deep-level understanding of why they work as well as they do, and in fact different researchers choose to emphasize rather different aspects of the peripheral auditory system's response to sound in their work. Most auditory models include a set of linear bandpass filters with bandwidth that increases nonlinearly with center frequency, a nonlinear rectification stage that frequently includes short-term adaptation and lateral suppression, and, in some cases, a more central display based on short-term temporal information. We estimate that the number of arithmetic operations of some of the currently-popular auditory models ranges from 35 to 600 times the number of operations required for the LPC-based processing used in SPHINX-II.

Pilot evalution of the Seneff auditory model. We recently completed a series of pilot evaluations using an implementation of the Seneff auditory model [21] on the census databse. Since almost all evaluations of physiologically-motivated front ends to date have been performed using artifically-added white Gaussian noise, we have been interested in the extent to which auditory models can provide useful improvements in recognition accuracy for speech that has been degraded by reverberation or other types of linear filtering. As in the case of microphone arrays, we are also especially interested in determining the extent to which improvements in robust-

ness provided by auditory modelling complement those that we already enjoy by the use of acoustical pre-processing algorithms such as CDCN.

We compared error rates obtained using the standard 12 LPC-based cepstral coefficents normally input to the SPHINX system, with those obtained using an implementation of the 40-channel mean-rate output of the Seneff model [21], and with the 40-channel outputs of Seneff's Generalized Synchrony Detectors (GSDs). The system was evaluated using the original testing database from the census task with the CLSTLK and PZM6FS microphones, and also with white Gaussian noise artificially added at signal-to-noise ratios of +10, +20, and +30 dB, measured using the global SNR method described in [19].

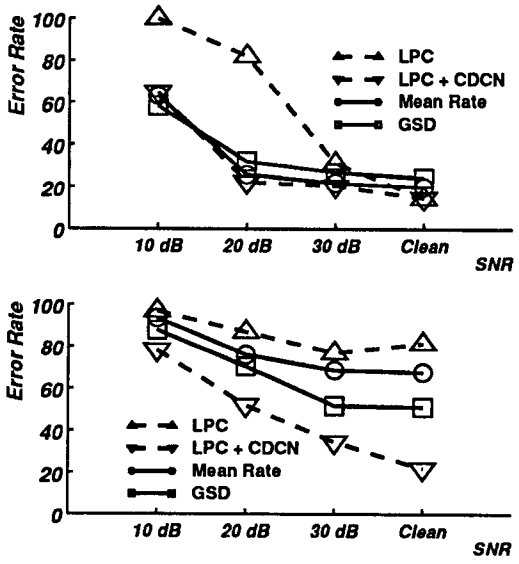

Figure 4: Pilot data comparing error rates obtained on the census task using the conventional LPC-based processing of SPHINX with results obtained using the mean rate and synchrony outputs of the Seneff auditory model. SPHINX was trained on the CLSTLK microphone in all cases, and tested using either the CLSTLK microphone (upper panel) or the Crown PZM6FS microphone (lower panel). White noise was artificially added to the speech signals and data are plotted as a function of global SNR.

Figure 4 summarizes the results of these comparisons, with error rate plotted as a function of SNR using each of the three peripheral signal processing schemes. The upper panel describes recognition error rates obtained with the system both trained and tested using the CLSTLK microphone, and the lower panel describes error rates obtained with the system trained with the CLSTLK microphone but tested with the PZM6FS microphone. When the system is trained and tested using the CLSTLK microphone, best performance is obtained using conventional LPC-based signal processing for "clean" speech. As the SNR is decreased, however, error rates obtained using either the mean rate or GSD outputs of the Seneff model degrade more gradually confirming similar findings from previous studies. The results in the lower panel of Fig. 4, demonstrate that the mean rate and GSD outputs of the Seneff model provide lower error rates than conventional LPC cepstra when the system is trained using the CLSTLK microphone and tested using the PZM6FS. Nevertheless, the level of performance achieved by the present implementation of the auditory model is not as good as that achieved by conventional LPC cepstra combined with the CDCN algorithm on the same data (Fig. 1). Furthermore, the combination of conventional LPC-based processing and the CDCN algorithm produced performance that equaled or bettered the best performance obtained with the auditory model for each test condition. Because the auditory model is nonlinear and not easy to port from one site to another, these comparisons should all be regarded as preliminary. It is quite possible that performance using the auditory model could further improve if greater attention were paid to tuning it to more closely match the characteristics of SPHINX.

We also attempted to determine the extent to which a combination of auditory processing and the CDCN algorithm could provide greater recognition accuracy than either processing scheme used in isolation. In these experiments we combined the effects of CDCN and auditory processing by resynthesizing the speech waveform from cepstral coefficients that were produced by the original LPC front end and then modified by the CDCN algorithm. The resynthesized speech, which was totally intelligible, was then passed through the Seneff auditory model in the usual fashion. Unfortunately, it was found that this particular combination of CDCN and the auditory model did not improve the recognition error rate beyond the level achieved by CDCN alone. A subsequent error analysis revealed that this concatenation of cepstral processing and the CDCN algorithm, followed by resynthesis and processing by the original SPHINX front end, degraded the error rates even in the absence of the auditory processing, although analysis and resynthesis without the CDCN algorithm did not produce much degradation. This indicates that useful information for speech recognition is lost when the resynthesis process is performed after the CDCN algorithm is run. Hence we regard this experiment as inconclusive, and we intend to explore other types of combinations of acoustical pre-processing with auditory modelling in the future.

5. SUMMARY AND CONCLUSIONS

In this paper we describe our current research in acoustical pre-processing for robust speech recognition, as well as our first attempts to integrate pre-processing with other approaches to robust speech recognition. The CDCN algorithm was also applied to the ATIS task for the first time, and provided the best recognition scores for speech collected using the unidirectional desktop PCC160 microphone. We demonstrated that the CDCN algorithm and the Flanagan delay-and-sum microphone array can provide complementary benefits to speech recognition in reverberant environments. We also found that the Seneff auditory model improves recognition accuracy of the CMU speech system in reverberant as well as noisy environ-

ments, but preliminary efforts to combine the auditory model with the CDCN algorithm were inconclusive.

ACKNOWLEDGMENTS

This research is sponsored by the Defense Advanced Research Projects Agency, DoD, through ARPA Order 7239, and monitored by the Space and Naval Warfare Systems Command under contract N00039-91-C-0158. Views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or of the United States Government. We thank Hsiao-Wuen Hon, Xuedong Huang, Kai-Fu Lee, Raj Reddy, Eric Thayer, Bob Weide, and the rest of the speech group for their contributions to this work. We also thank Jim Flanagan, Joe French, and A. C. Surendran for their assistance in obtaining the experimental data using the array microphone, and Stephanie Seneff for providing source code for her auditory model. The graduate studies of Tom Sullivan and Yoshiaki Ohshima have been supported by Motorola and IBM Japan, respectively.

REFERENCES

1. Juang, B. H., "Speech Recognition in Adverse Envoronments", *Comp. Speech and Lang.*, Vol. 5, 1991, pp. 275-294.

2. Acero, A. and Stern, R. M., "Environmental Robustness in Automatic Speech Recognition", *ICASSP-90*, April 1990, pp. 849-852.

3. Erell, A. and Weintraub, M., "Estimation Using Log-Spectral-Distance Criterion for Noise-Robust Speech Recognition", *ICASSP-90*, April 1990, pp. 853-856.

4. Acero, A. and Stern, R. M., "Robust Speech Recognition by Normalization of the Acoustic Space", *ICASSP-91*, May 1991, pp. 893-896.

5. Liu, F.-H., Acero, A., and Stern, R. M., "Efficient Joint Compensation of Speech for the Effects of Additive Noise and Linear Filtering", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 1992.

6. Stockham, T. G., Cannon, T. M,. and Ingebretsen, R. B. , "Blind Deconvolution Through Digital Signal Processing", *Proc. IEEE*, Vol. 63, 1975, pp. 678-692.

7. Boll, S. F. , "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", *ASSP*, Vol. 27, 1979, pp. 113-120.

8. Sakoe, H., Chiba, S., "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 26, 1978, pp. 43-49.

9. Hermansky, H., Morgan, N., Bayya, A., Kohn, P., "Compensation for the Effect of the Communication Channel in Auditory-Like Analysis of Speech (RASTA-PLP)", *Proc. of the Second European Conf. on Speech Comm. and Tech.*, September 1991.

10. Ward, W., Issar, S., Huang, X., Hon, H.-W., Hwang, M.-Y., Young, S., Matessa, M., Liu, F.-H., Stern, R., "Speech Understanding in Open Tasks", *Proc. of the DARPA Speech and Natural Language Workshop*, February 1992.

11. Flanagan, J. L., Johnston, J. D., Zahn, R., and Elko, G.W., "Computer-steered Microphone Arrays for Sound Transduction in Large Rooms", *The Journal of the Acoustical Society of America*, Vol. 78, Nov. 1985, pp. 1508-1518.

12. Flanagan, J. L., Berkeley, D. A., Elko, G. W., West, J. E., and Sondhi, M. M., "Autodirective microphone systems", *Acustica*, Vol. 73, February 1991, pp. 58-71.

13. Widrow, B., and Stearns, S. D., *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

14. Peterson, P. M., "Adaptive Array Processing for Multiple Microphone Hearing Aids". RLE TR No. 541, Res. Lab. of Electronics, MIT, Cambridge, MA

15. Alvarado, V. M., Silverman, H. F., "Experimental Results Showing the Effects of Optimal Spacing Between Elements of a Linear Microphone Array", *ICASSP-90*, April 1990, pp. 837-840.

16. Van Compernolle, D., "Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1990, pp. 833-836.

17. Lyon, R. F., "A Computational Model of Binaural Localization and Separation", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1983, pp. 1148-1151.

18. Cohen, J. R., "Application of an Auditory Model to Speech Recognition", *The Journal of the Acoustical Society of America*, Vol. 85, No. 6, June 1989, pp. 2623-2629.

19. Ghitza, O., "Auditory Nerve Representation as a Front-End for Speech Recognition in a Noisy Environment", *Comp. Speech and Lang.*, Vol. 1, 1986, pp. 109-130.

20. Lyon, R. F., "A Computational Model of Filtering, Detection, and Compression in the Cochlea", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 1982, pp. 1282-1285.

21. Seneff, S., "A Joint Synchrony/Mean-Rate Model of Auditory Speech Processing", *Journal of Phonetics*, Vol. 16, No. 1, January 1988, pp. 55-76.

22. Hunt, M., "A Comparison of Several Acoustic Representations for Speech Recognition with Degraded and Undegraded Speech", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 1989.

23. Meng, H., and Zue, V. W., "A Comparative Study of Acoustic Representations of Speech for Vowel Classification Using Multi-Layer Perceptrons", *Int. Conf. on Spoken Lang. Processing*, November 1990, pp. 1053-1056.